Final Technical Report

USGS Award Number:	G20AP00019
Title of Award:	Detection of Ground Motion Spatial Correlation Nonstationarities Using Network Analysis
Authors:	Jack Baker Stanford University Yang & Yamasaki Building, Room 283 423 Via Ortega Stanford, CA 94305-4020 650-725-2573 (telephone) 650-723-7514 (fax) bakerjw@stanford.edu Yilin Chen Stanford University 439 Panama Mall Building 02-540 Stanford CA 94305 yilinc2@stanford.edu
Term Covered by the Award:	1/1/2020 to 12/31/2020

Acknowledgement of Support

This work was supported by the U.S. Geological Survey (USGS) via External Research Program award G20AP00019.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey.

Abstract

In this report, we propose a community detection method to find regions in spatial data with higher correlations. We construct a 'correlation deviation graph' that considers the influence of global correlation caused by the node's geographical location. We compare the performance of the correlation deviation graph to conventional correlation graph construction approaches for community detection, using synthetic data, and show that the proposed method has the best performance in the presence of a global correlation structure. We also compare the performance of signed spectral clustering and the signed Louvain algorithm for community detection on the correlation graphs, and find that the signed spectral clustering algorithm performs best. We apply our algorithms to simulated earthquake ground motion data in the Los Angeles region. The results suggest that communities of high correlation in ground shaking tend to be associated with common geological conditions and relative location along the rupture strike direction.

1 Introduction

Spatial data is measured data with an associated geographical location. A common property of spatial data is that spatially closer sites have higher correlation, which is usually referred to as spatial correlation. These correlations have been studied for various geoscience data, such as earthquake shaking, wind speed, and temperature (e.g., Alexiadis *et al.*, 1999; Boore *et al.*, 2003; North *et al.*, 2011). Modeling spatial correlation in earthquake shaking is essential for regional seismic risk analysis (Park *et al.*, 2007; Goda and Hong, 2008; Weatherill *et al.*, 2015). For example, when studying the risk of distributed infrastructure, higher spatial correlations in shaking produce larger tail risk (Jayaram and Baker, 2010).

Traditionally, spatial correlation is studied using geostatistics methods such as semivariograms (Goovaerts, 1997). However, these approaches usually assume stationarity (i.e., correlation depends only on distance), which results in a global correlation model. Alternatively, the spatial correlation can be modeled using a graph, where nodes are embedded in space, and edges describe their correlation. A positive edge indicates that two nodes are positively correlated and vice versa. One of the essential features in graphs is the existence of community structure: nodes of a graph can be grouped into communities based on their similarity. The procedure of identifying such groups is referred to as community detection. Community detection algorithms have been widely used in various fields, including user recommendations in social networks (e.g., Kanavos *et al.*, 2018), drug discovery in protein interaction networks (e.g., Bhowmick and Seah, 2015) and analysis of passenger travel patterns in transportation networks (e.g., Yildirimoglu and Kim, 2018).

Communities in correlation graphs have been widely studied in the correlation clustering field (Emanuel and Fiat, 2003; Bansal *et al.*, 2004; Demaine *et al.*, 2006). The objective of correlation clustering is to identify groups of nodes as communities, such that nodes within communities are more correlated. Various methods have been proposed to solve this problem. Generally speaking, they express the objectives as mathematical terms and try to optimize it using a search procedure (Duan *et al.*, 2014). Among these, spectral-based (Ng *et al.*, 2002; Von Luxburg, 2007) and modularity-based methods (Clauset *et al.*, 2004; Blondel *et al.*, 2008) are commonly used. However, when we use the graph to model spatial variables, the topology alone does not contain all the information (Barthélemy, 2011; Leskovec and Horvitz, 2014), because the locations of nodes matter. Therefore, when we analyze the graph with spatial location information, we have to consider the physical distance of nodes.

Significant research has been done to find communities in a spatial network. Existing methods generally utilize location information in community detection. In this context, existing community detection algorithms aim to find a group of nodes that are both tightly connected in a graph and located in a similar geographic

region. Typically, the methods include density-based community approaches (Yao *et al.*, 2019; Wu *et al.*, 2017; Ester *et al.*, 1996), adding constraints to detected communities (Chen *et al.*, 2015), or encoding location information during community detection (Van Gennip *et al.*, 2013; Yiu and Mamoulis, 2004). Applications include clustering on transportation networks and discovering Geo-Social relations. However, the formulation of problems for these spatial networks is different from earthquake instrument networks. This is because the locations of instruments may influence the recorded ground motions, so the influence of spatial location should be removed in order to find the underlying signal. Similar work has been done by Expert *et al.* (2011), where a distance deterrence function is used to remove the effect of space in order to reveal hidden structural similarities between the nodes. However, limited studies have considered earthquake instrument networks.

In this report, we propose a method to construct a 'correlation deviation graph' for detecting communities of nodes with higher within-community correlations than would be expected when considering only their physical distance. We compare it to two traditional graph construction methods (the unweighted correlation graph and Pearson's correlation graph) for two community detection algorithms (signed Louvain and signed spectral clustering) on synthetic and real-world data.

Section 2 of the report describes existing methods for representing correlation data in graphs, and detecting communities in these graphs. Section 3 describes the method proposed in this report. Then Section 4 evaluates the performance of the existing and proposed methods on example synthetic data with known correlation structure. After seeing that the proposed method works well on synthetic data, Section 5 applies the method to data from simulated earthquake ground motions.

2 Related work

2.1 Correlation Graph

There are various ways to represent a correlation graph. In this section, we review two common methods, Pearson's correlation graph, and the unweighted correlation graph.

2.1.1 Pearson's correlation graph

The simplest way to construct a correlation graph is to use correlation coefficients as links to represent the correlation level between nodes. Given $X = [x_1, x_2, ..., x_n]$, representing m observations at n nodes, where x_i is the observation vector of length m at node i, the Pearson's correlation graph can be constructed by computing correlation coefficients for observations between every pair of nodes:

$$\rho(i,j) = \frac{(x_i - \bar{x}_i)^T (x_j - \bar{x}_j)}{||x_i - \bar{x}_i||\,||x_j - \bar{x}_j||} \tag{1}$$

where $\bar{x}_i$ is the mean of the observation vector at node i and $||x_i - \bar{x}_i||$ is the l^2 norm of the centered observation vector. Then the constructed Pearson's correlation graph consists of $\rho(i,j)$ of all pairs of nodes as edges. This results in a complete graph with signed and weighted links. Pearson's correlation graphs have been widely used for data mining tasks, such as evaluating behaviors of interestingness measures (Huynh *et al.*, 2006) and unsupervised manifold learning in image retrieval tasks (Pedronette and Torres, 2016).

However, there are several limitations for Pearson's correlation graph. First, using correlation coefficients for all pairs of nodes results in a complete graph with $\frac{n(n-1)}{2}$ edges, which in some cases will be too large to analyze. Second, negative links can exist in the graph, which limits the use of some community detection algorithms. Third, there is considerable uncertainty in correlation coefficient estimation, especially when the underlying correlation structure is weak and the number of observations used to calculate correlation coefficients is small, which may hinder detection of community structure.

2.1.2 Unweighted correlation graph

Another way to find correlated communities is to construct a graph with only highly correlated nodes, by setting a threshold θ, and discarding all links with correlation lower than this threshold (MacMahon and Garlaschelli, 2015). The adjacency matrix A of an unweighted correlation graph is

$$A_{ij} = \begin{cases} 1 & \rho(i,j) > \theta \\ 0 & \rho(i,j) \leq \theta \end{cases} \tag{2}$$

where A_{ij} is the edge weight between nodes i and j, with $A_{ij} = 1$ if node i and node j are connected and $A_{ij} = 0$ if not. The remaining disconnected sets of nodes can be inferred as communities. The unweighted correlation graph is commonly used in finance fields (e.g., asset graph). This method is robust to uncertainty in correlation coefficient estimation, but has several other drawbacks. First, it ignores low-correlated and negatively correlated links that can be useful in practice to identify communities (Esmailian and Jalili, 2015). Second, the results highly depend on the threshold choice. Finally, a global threshold for all nodes may not be appropriate. For example, if a set of nodes are internally more correlated but less correlated with outside nodes, it will fail to discover this internal community if the internal correlation is less than the threshold.

2.2 Community Detection Algorithms on Correlation Graphs

This section discusses two algorithms for detecting communities in signed and weighted graphs such as the correlation graphs used in this study.

2.2.1 Signed Spectral Clustering

Kunegis *et al.* (2010) proposes a signed version of spectral clustering by defining a signed weighted adjacency matrix A, where A_{ij} is the edge weight between nodes i and j. The signed degree matrix is defined as:

$$\tilde{D}_{ii} = \sum_i |A_{ij}| \tag{3}$$

$\tilde{D}$ is a diagonal matrix where the value $\tilde{D}_{ii}$ is the sum of absolute value of edge weights that connect to node i. The signed Laplacian matrix is then defined as:

$$\tilde{L} = \tilde{D} - A \tag{4}$$

and the signed ratio cut between community X and Y is

$$\text{SignedRatioCut} = \text{scut}(X,Y)\left(\frac{1}{|X|} + \frac{1}{|Y|}\right) \tag{5}$$

where $|X|$ and $|Y|$ are the sizes of the communities, and $\text{scut}(X,Y)$ can be calculated as:

$$\text{scut}(X,Y) = 2\text{cut}^+(X,Y) + \text{cut}^-(X,X) + \text{cut}^-(Y,Y) \tag{6}$$

$$\text{cut}^+(X,Y) = \sum_{i\in X, j\in Y} A^+_{ij} \tag{7}$$

$$\text{cut}^-(X,Y) = \sum_{i\in X, j\in Y} A^-_{ij} \tag{8}$$

$$A^+_{ij} = \max(0, A_{ij}) \tag{9}$$

$$A^-_{ij} = \max(0, -A_{ij}) \tag{10}$$

where A_{ij}^{+} is the adjacency matrix with only positive edges, while A_{ij}^{-} is the adjacency matrix with only negative edges. The signed cut, $\text{scut}(X,Y)$, counts the number of positive edges that connect X to Y ($\text{cut}^{+}(X,Y)$) and the number of negative edges that remain within these groups ($\text{cut}^{-}(X,X)+\text{cut}^{-}(Y,Y)$). Minimizing the signed ratio cut is equivalent to finding a set of disconnected communities where the nodes within communities are strongly correlated, but nodes across communities are less correlated. Minimizing the signed ratio cut $\text{scut}(X,Y)$ can be approximately solved by using eigenvectors of $\tilde{L}$ to partition the graph into groups (Kunegis *et al.*, 2010). In this report, we use K-Means clustering on these eigenvectors to partition the graph into communities.

2.2.2 Signed Louvain Algorithm

A second community detection approach is the signed Louvain algorithm to compute graph modularity (Blondel *et al.*, 2008; Gómez *et al.*, 2009). The signed graph modularity is defined as:

$$\tilde{Q} = \frac{1}{2w^{+}+2w^{-}} \sum_{i}\sum_{j} \delta(C_i, C_j) \left[w_{ij} - \left(\frac{w_i^{+} w_j^{+}}{2w^{+}} - \frac{w_i^{-} w_j^{-}}{2w^{-}} \right) \right] \tag{11}$$

where C_i is the community to which node i is assigned, $\delta(C_i, C_j)$ takes value 1 if $C_i = C_j$ and 0 otherwise, and

$$w_{ij} = A_{ij}^{+} - A_{ij}^{-} \tag{12}$$

$$w_i^{+} = \sum_{j} A_{ij}^{+} \tag{13}$$

$$w_i^{-} = \sum_{j} A_{ij}^{-} \tag{14}$$

$$w^{+} = \sum_{i}\sum_{j} A_{ij}^{+} \tag{15}$$

$$w^{-} = \sum_{i}\sum_{j} A_{ij}^{-} \tag{16}$$

where A_{ij}^{+} and A_{ij}^{-} are the signed adjacency matrices defined in Equation 9 and 10. To optimize the modularity, the modularity gain of assigning node i to community C can be calculated as:

$$\Delta\tilde{Q}(i \to C) = \frac{2w^{+}}{2w^{+}+2w^{-}} \Delta\tilde{Q}^{+} - \frac{2w^{-}}{2w^{+}+2w^{-}} \Delta\tilde{Q}^{-} \tag{17}$$

where $\Delta\tilde{Q}^{+}$ is the positive modularity gain and $\Delta\tilde{Q}^{-}$ is the negative modularity gain, which can be calculated as:

$$\begin{aligned} \Delta\tilde{Q}^{+} = &\left[\frac{\sum_{in^{+}} + k_{i,in}^{+}}{2w^{+}} - \left(\frac{\sum_{tot^{+}} + k_i^{+}}{2w^{+}} \right)^2 \right] \\ &- \left[\frac{\sum_{in^{+}}}{2w^{+}} - \left(\frac{\sum_{tot^{+}}}{2w^{+}} \right)^2 - \left(\frac{k_i^{+}}{2w^{+}} \right)^2 \right] \end{aligned} \tag{18}$$

$$\begin{aligned} \Delta\tilde{Q}^{-} = &\left[\frac{\sum_{in^{-}} + k_{i,in}^{-}}{2w^{-}} - \left(\frac{\sum_{tot^{-}} + k_i^{-}}{2w^{-}} \right)^2 \right] \\ &- \left[\frac{\sum_{in^{-}}}{2w^{-}} - \left(\frac{\sum_{tot^{-}}}{2w^{-}} \right)^2 - \left(\frac{k_i^{-}}{2w^{-}} \right)^2 \right] \end{aligned} \tag{19}$$

where $k^+_{i,in}$ and $k^-_{i,in}$ are the sums of positive/negative weights between node i and C, k^+_i and k^-_i are the sums of all positive/negative link weights of node i, $\sum_{in^+}$ and $\sum_{in^-}$ are the sums of positive/negative link weights between nodes in C, and $\sum_{tot^+}$ and $\sum_{tot^-}$ are the sums of all positive/negative link weights of nodes in C. This method iteratively assigns a node i to a community C with the highest $\Delta\tilde{Q}(i \to C)$ until it reaches a maximum.

The two community detection algorithms described above both try to maximize the positive edge weights within communities, while minimizing the negative edge weights within communities and positive edge weights across communities. Therefore, if the edge weights represent the correlation level between nodes, the nodes in a community will have higher correlations, and nodes in different communities will have lower correlations.

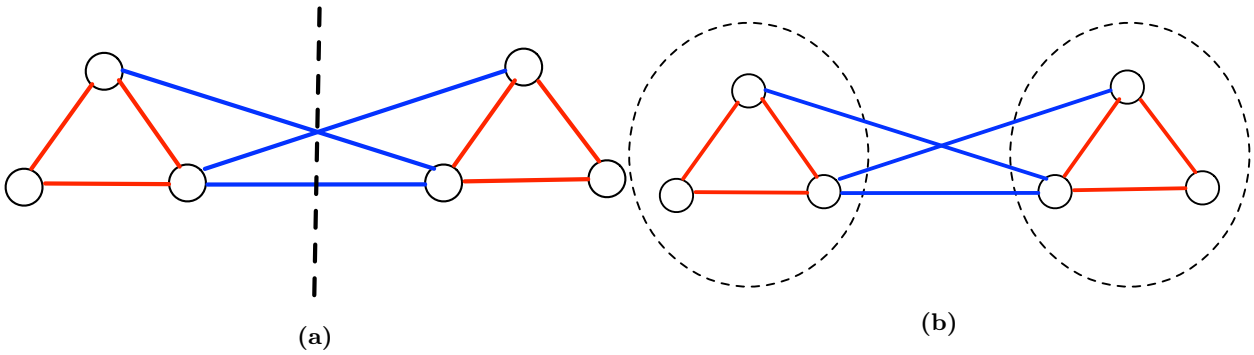


Figure 1: Illustration of the two community detection algorithms with two well-defined communities in a graph. Red lines correspond to positive links, and blue lines correspond to negative links. (a) The dashed line shows the optimal graph cut by the signed spectral clustering algorithm. (b) The dashed lines show the node group assignment with the highest modularity score by the signed Louvain algorithm.

3 Correlation Deviation Graph

Although the above correlation graphs and community detection algorithms can be applied to spatial graphs (i.e., nodes with geographical locations), they do not account for the fact that nodes that are spatially closer to each other generally have a larger correlation. Additionally, they do not recognize that a correlation coefficient estimate is more statistically reliable if there are more observations. Here we propose a method to construct a correlation deviation graph that considers the influences of geographical locations and sample variance of correlation estimation. In the remainder of this section, nodes with geographical locations are called stations.

3.1 Problem Statement

We have n stations with separation distances $d(i, j)$, where $d(i, j)$ is the distance between station i and j, and a data matrix $X = [x_1, x_2, ..., x_n]$ of m observations at n stations. We want to construct a graph such that the weight of the edge represents how much a pair of stations' correlation is lower or higher than expected based on their separation distances.

3.2 Methodology

1. Calculate the pairwise correlation coefficient $\rho(i, j)$ for all pairs of stations i and j using equation 1.

2. Fit a global correlation function

$$\hat{\rho}(i,j) = f(d(i,j)) \tag{20}$$

The global correlation function describes the expected correlation of two nodes caused by their distance. In general, the correlation will decay as distance increases. The function must produce a positive definite correlation matrix for the graph (Goovaerts, 1997).

3. Construct a correlation deviation graph with adjacency matrix

$$A_{ij} = \left[Z\left(\rho(i,j)\right) - Z\left(\hat{\rho}(i,j)\right)\right] \times \sqrt{(m-3)} \tag{21}$$

where m is the number of observations used to calculate the $\rho(i,j)$, and $Z(x) = \ln \frac{1+x}{1-x}$ is the Fisher transformation. It transforms the sample correlation coefficient $\rho(i,j)$ to a normally distributed random variable, and the standard deviation of $Z(\rho(i,j))$ is $\frac{1}{\sqrt{m-3}}$. Therefore, if $\hat{\rho}(i,j)$ is the expected correlation coefficient, A_{ij} will follow the standard normal distribution, which quantify the correlation deviation of a pair of station relative to the expected correlation.

4. Run community detection on the signed weighted graph with stations as nodes and weight of links from the adjacency matrix A defined in Equation 21. The objective of community detection is to jointly maximize the positive edge weights within communities and minimize the negative edge weights within communities.

4 Synthetic Data Example

In this section, we show how a global correlation can influence community detection on a correlation graph. We illustrate this first using the simplest case that the global correlation is constant between all nodes, and then a more realistic example that the global correlation is a function of the distance between the nodes.

4.1 Data Generation

Here we consider synthetic correlation data from a correlation matrix Σ with entries defined as follows

$$\Sigma_{ij} = \begin{cases} 1 & i = j \\ \rho_0(i,j) & i \neq j,\ C_i \neq C_j \\ \rho_0(i,j) + \Delta & i \neq j, C_i = C_j \end{cases} \tag{22}$$

where Σ_{ij} represents the correlation between the observations at node i and node j. There is a global correlation value $\rho_0(i,j)$ for all pairs of nodes in the graph, plus a correlation from the community structure: the correlation is higher by Δ if the nodes are in the same community. We consider two types of global correlation functions:

$$\rho_0(i,j) = \begin{cases} c & \textit{example 1} \\ e^{-0.3d(i,j)} & \textit{example 2} \end{cases} \tag{23}$$

For the first example, the global correlation is constant (i.e., $\rho_0(i,j) = c$ for all node pairs). For the second example, we assume an exponentially decaying global correlation function $\rho_0(i,j) = e^{-0.3d(i,j)}$, where $d(i,j)$ is the distance between node i and j.

To test the algorithms, observations at all nodes are sampled m times from a multivariate normal distribution with the specified correlation matrix, standard deviations of one, and means of zero, so we get a data

matrix of m observations at all nodes $X = [x_1, x_2, ..., x_n]$. We compute correlation coefficients for each pair of nodes and then construct the Pearson's correlation graph, unweighted correlation graph, and correlation deviation graph. The expected correlation function of the correlation deviation graph and the threshold of the unweighted correlation graph are set to $\rho_0(i, j)$.

4.2 Evaluation

We next detect highly correlated communities by partitioning the graph into groups. Since we know the true communities of nodes from which the data were sampled, the external clustering evaluation $F-$Score is used (Powers, 2011):

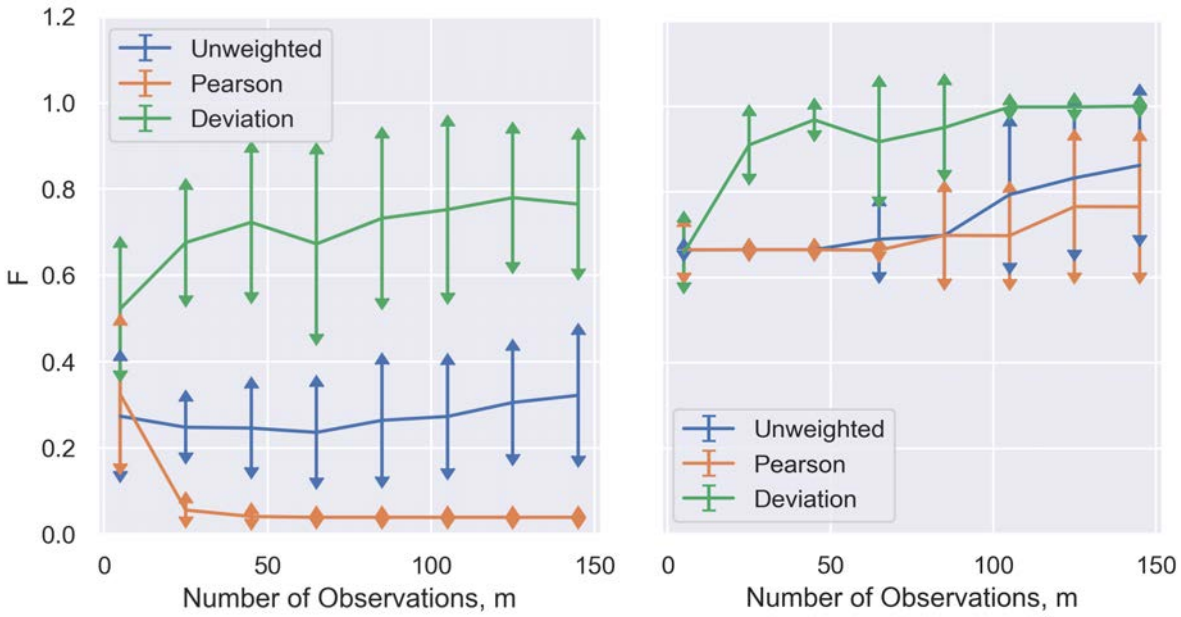
$$F = \sum_j p_j \max_i \left[2\frac{p_{ij}}{p_i}\frac{p_{ij}}{p_j} / (\frac{p_{ij}}{p_i} + \frac{p_{ij}}{p_j}) \right] \tag{24}$$

where $p_{ij} = \frac{n_{ij}}{n}$, $p_i = \frac{n_{i.}}{n}$, $p_j = \frac{n_{j.}}{n}$, and n_{ij} is the number of nodes from partition group i that are in community j, $n_{i.}$ is the number of nodes in partition i and $n_{j.}$ is the number of nodes in community j. Bad community assignments have F-scores close to 0, and a perfect community assignment has an F-score of 1.

4.3 Example 1, Constant Global Correlation

In this example, we consider a constant global correlation with two equal-size communities and 100 nodes ($n = 100$). We evaluate the performance of the correlation deviation graph, unweighted correlation graph, and Pearson's correlation graph by running community detection algorithms on these graphs and comparing the detected communities with ground-truth labels using the $F-$Score.

This sampling exercise is repeated ten times, and the samples are used to get the mean and standard deviation of the $F-$Scores. Then the calculation is repeated for varying numbers of observational data, and varying values of the model parameters, to quantify the performance of each algorithm over a range of conditions.

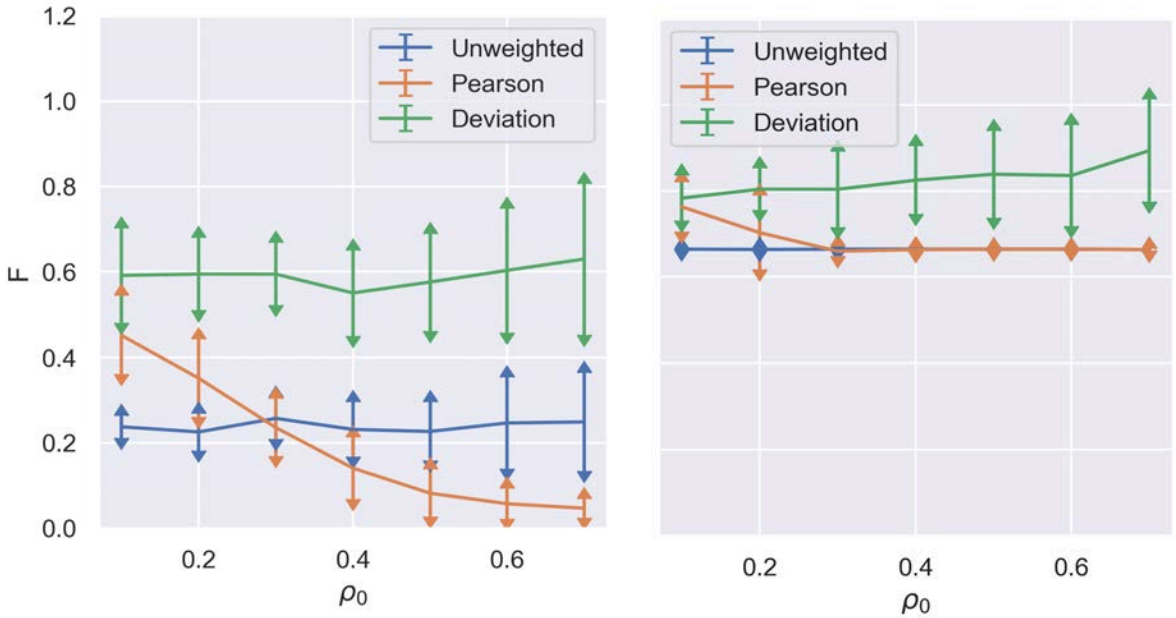


(a) **(b)**

Figure 2: Example 1: Clustering evaluation $F-$Scores using the unweighted correlation graph, Pearson's correlation graph, and correlation deviation graph, with (a) the Louvain algorithm and (b) spectral clustering. For all cases, $\rho_0 = 0.5$, $\Delta = 0.1$, and the number of observations is varied on the x axis. Lines show the mean values, and arrows show the $\pm$ standard deviations, of the $F-$Score results from ten samples.

Figure 2a shows the $F-$scores for three graph construction methods and varying numbers of observations m, using the Louvain algorithm. The correlation deviation graph has the highest accuracy in all cases. Its superiority is less significant when the number of observations is small and more significant when the number of observations is large. However, it cannot perfectly recover the true community structure even when the number of observations is greater than 100, as the F-score never reaches 1. The unweighted correlation graph does not converge to higher accuracy as the number of observations increases. This is because the edge weights in the graph are not scaled when we have more observations. Pearson's correlation graph fails to detect communities in this task, as it ignores the expected correlation information.

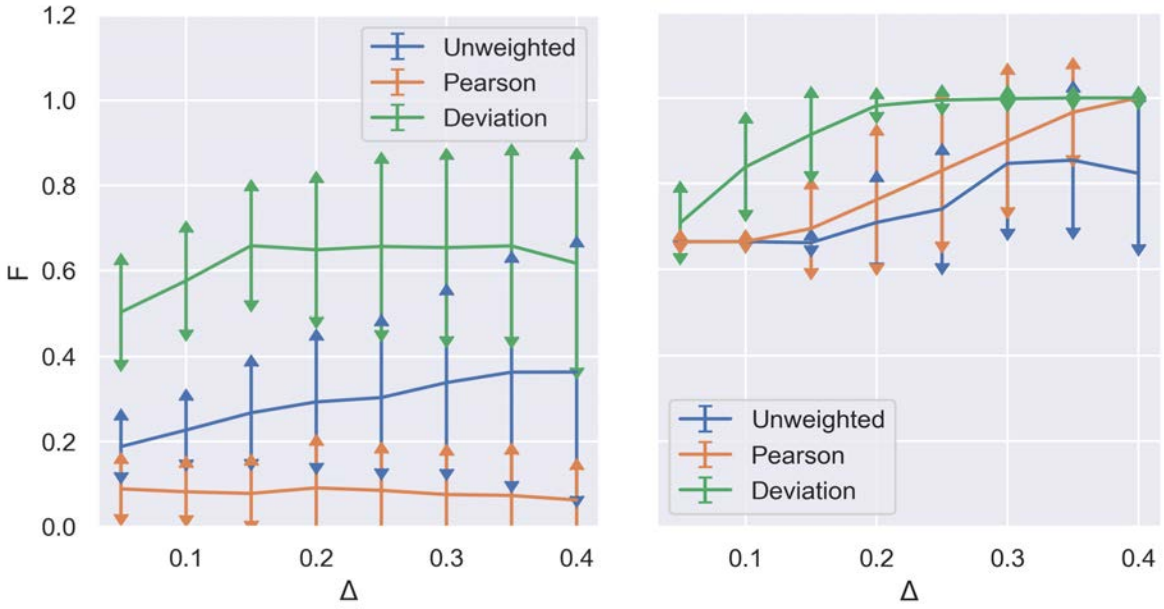
Figure 2b shows $F-$scores for the three graph construction methods, when using the spectral clustering algorithm to detect communities. The correlation deviation graph still has the highest accuracy in all stages. It can reliably detect the communities when $m > 20$. Increasing the number of observations increases the $F-$scores of the other two methods, but it is less significant than the correlation deviation graph method. This suggests that the correlation deviation graph is more robust for different clustering algorithms, and it can recover the community structure as the number of observations increases.



(a) **(b)**

Figure 3: Example 1: $F-$Scores of unweighted correlation graph, Pearson's correlation graph, and correlation deviation graph, with $m = 10$, $\Delta = 0.1$, and varying ρ_0 using (a) the Louvain algorithm and (b) spectral clustering. Lines show the mean values, and arrows show the $\pm$ standard deviations, of the $F-$Scores from ten samples.

We next vary ρ_0 to study its influence on community detection accuracy. In Figure 3, the correlation deviation graph still performs best for all ρ_0, and when using either the Louvain or spectral clustering algorithms. With Pearson's correlation graph, increasing ρ_0 will strongly decrease the detection accuracy, as with this approach the community structure tends to be hidden if the global correlation is substantial, especially when using the Louvain algorithm. The unweighted correlation graph method is less influenced by the global correlation ρ_0, but it performs worse than the correlation deviation graph.



(a) Louvain

(b) Spectral

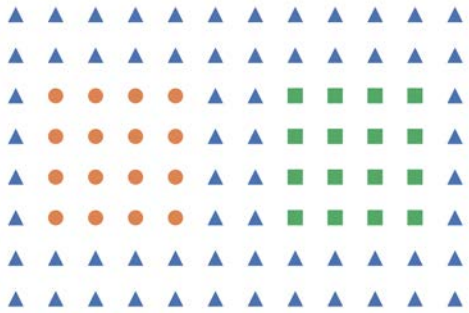
Figure 4: Example 1: $F-$Scores of unweighted correlation graph, Pearson's correlation graph, and correlation deviation graph, with $m = 10$, $\rho_0 = 0.5$, and varying Δ using (a) the Louvain algorithm and (b) spectral clustering. Lines show the mean values, and arrows show the $\pm$ standard deviations, of the $F-$Scores from ten samples.

Figure 4 shows the performance of the graph construction methods for varying Δ (where a large Δ means a stronger community structure). When using the Louvain algorithm, the correlation deviation graph performs significantly better than the others. For the spectral clustering algorithm, the correlation deviation graph performs slightly better than others when the community structure is weak (small Δ). The correlation deviation graph and unweighted correlation graph both converge to $F = 1$ as Δ increases, but the unweighted correlation graph does not.

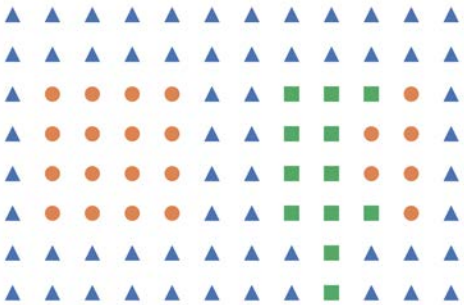
In summary, the correlation deviation graph outperforms others, especially for larger ρ_0 and lower numbers of observations. In general, spectral clustering performs better than the Louvain algorithm across different graph construction methods. The correlation deviation graph with spectral clustering produces results that converge to the true communities even when the underlying structure is hidden by a global correlation.

4.4 Example 2, Non-Constant Global Correlation

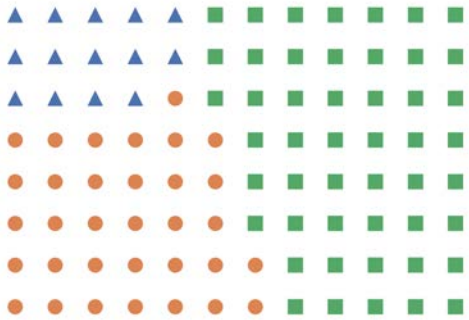
We now consider the exponentially decaying global correlation model. The nodes are embedded in a 2D grid, and the global correlation is a function of their spatial distance. Figure 5a shows the diagram of node locations and community assignments. There are three communities in the graph (indicated by three distinct symbols), and nodes within each community are set to have correlations that are $\Delta = 0.1$ higher than the global correlation. With regard to equation 23, $d(i, j) = 1$ corresponds to nodes that are adjacent (horizontally or vertically).



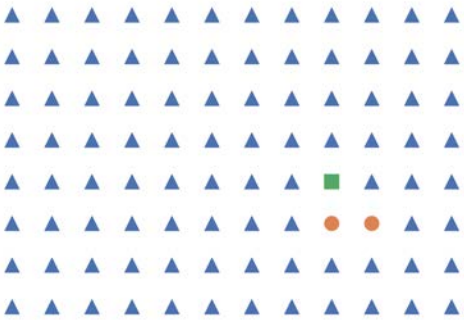
(a) True Communities



(b) Detected Communities on Deviation Graph, $F = 0.92$



(c) Detected Communities on Pearson Graph, $F = 0.56$



(d) Detected Communities on Unweighted Graph, $F = 0.63$

Figure 5: (a) True Communities. One realization of detected communities with m=145 observations, spectral clustering, and (b) correlation deviation graph, (c) Pearson's correlation graph, and (d) unweighted correlation graph. Nodes of different communities are shown in different symbols.

We simulated the data using the correlation matrix specified by the above community structure (with means=0 and standard deviations=1). We then constructed the three correlation graphs to be used with spectral clustering to find communities. Figure 5b shows the detected communities using the correlation deviation graph. It almost recovers the true community structure (with a few misclassified nodes). Figures 5c and 5d show the detected communities using the Pearson's correlation graph and unweighted correlation graph. Pearson's method identifies two major communities but fails to distinguish the two inside communities surrounded by the outside community. The unweighted correlation graph is unable to identify this type of community structure, as the global correlation cannot be distinguished from the within-community features.

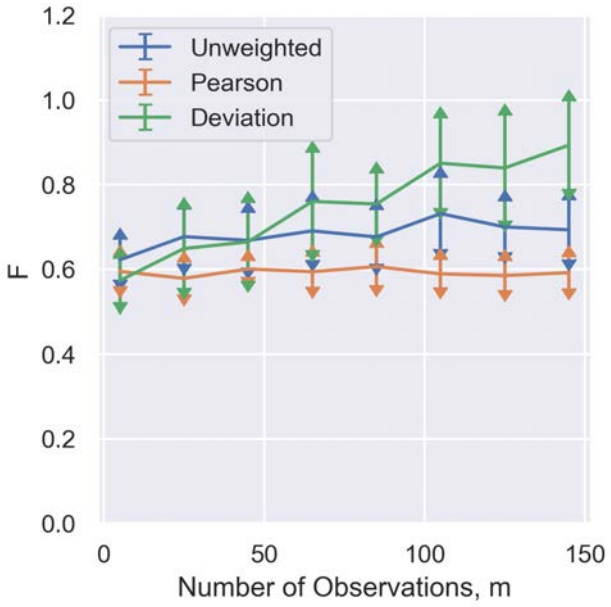


Figure 6: Example 2: $F-$Scores when using spectral clustering with the correlation deviation graph, Pearson's correlation graph and unweighted correlation graph, with exponentially decayed ρ_0, $\Delta = 0.1$, and varied numbers of observations. Lines show the mean values, and arrows show the $\pm$ standard deviations, of the $F-$Score results from ten samples.

Figure 6 shows the F scores of the three methods with varied numbers of observations. Compared with the first example, increasing the number of observations is less helpful for the Pearson's correlation graph and unweighted correlation graph. The correlation deviation graph method can still achieve high accuracy, given a sufficient number of observations.

In summary, the correlation deviation graph method still outperforms others in the more complex second community detection example. When the communities are embedded spatially, the Pearson's and unweighted correlation graph methods tend to make a coarse graph partition but fail to detect detailed communities. The correlation deviation graph method can recover the true communities, but it requires more observations than in the first example.

4.5 Optimal Number of Communities

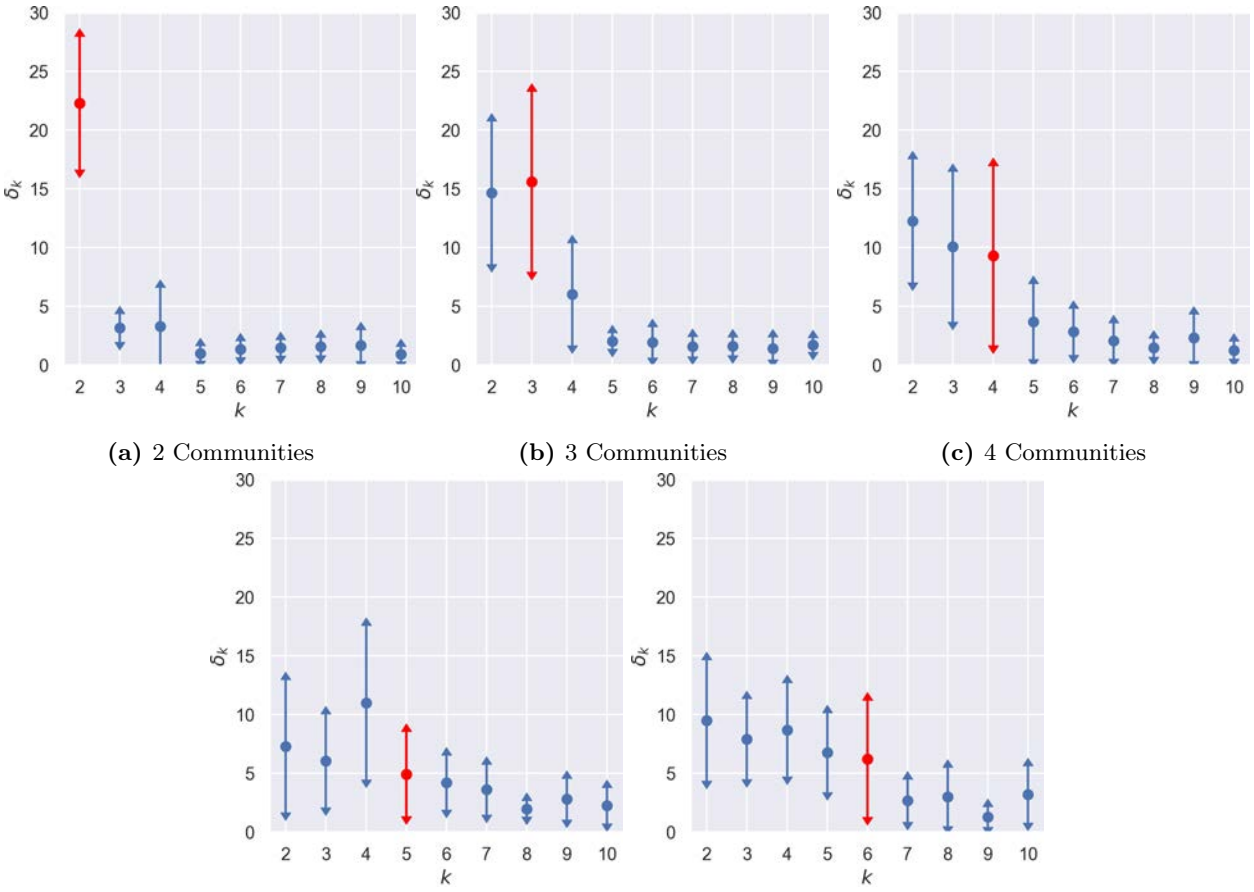
Although we have shown that spectral clustering on the correlation deviation graph can recover the true community structure, the number of communities is a user-defined parameter. In real-world examples with an unknown number of communities, the challenge of finding the optimal number of communities remains. In this section, we show that a gap statistic based on eigenvalues can be used as a heuristic to determine the optimal number of communities.

The difference between two consecutive eigenvalues, termed the eigengap, is defined as:

$$\delta_k = |\lambda_k - \lambda_{k+1}| \tag{25}$$

where λ_k is kth smallest eigenvalue of the signed Laplacian matrix $\tilde{L}$ in Equation 4. Eigengaps have been

used as a heuristic in previous studies to estimate the optimal number of communities in unsigned graphs (Zelnik-Manor and Perona, 2005; Ng *et al.*, 2002). We adopt the same approach here, except that the first eigenvalue is dropped because the smallest eigenvalue of $\tilde{L}$ is not equal to zero. This is because our graph is signed, and thus the row sum of $\tilde{L}$ will not equal to zero. We simulate correlation deviation graphs with various numbers of communities, using $m = 100$, $\Delta = 0.1$, and $\rho_0 = 0.5$, and compute their eigengaps.



(a) 2 Communities **(b)** 3 Communities **(c)** 4 Communities

(d) 5 Communities **(e)** 6 Communities

Figure 7: Eigengaps of the simulated correlation deviation graphs with various specified number of communities. Solid circles show the mean values, and arrows show the ± standard deviations, of the eigengaps computed from ten samples. The eigengap at the true number of communities is highlighted in red.

Figure 7 shows the first ten eigengaps computed from simulated graphs with a specified number of communities. The eigengap at the true number of communities is shown in red. For k = 2, 3, 4, 6, the eigengaps when k is less than the true number of communities are large, and they drop for larger k. This suggests that the position of the last significant eigengap is an indicator of the number of communities in the data.

5 Application: Correlated Earthquake Ground Shaking

We now apply the preferred correlation deviation graph method to earthquake ground motion data. Specifically, we are interested in the intensity of shaking observed during an earthquake. A number of models

for ground motion correlations have been calibrated using observational data, and in all of those models the correlation depends only on the separation distance between the stations of interest (e.g., Jayaram and Baker, 2009; Goda and Atkinson, 2010; Loth and Baker, 2013; Markhvida *et al.*, 2018; Heresi and Miranda, 2019). However, some researchers have studied the spatial non-stationarity and anisotropy of ground motion residuals (Sgobba *et al.*, 2019; Garakaninezhad and Bastami, 2017; Lanzano *et al.*, 2017), due to systematic source, path and site effects. Therefore, anticipation that a more complex correlation structure may actually exist in these data motivated our development of the proposed method.

5.1 Data

We consider earthquake ground motion data at 335 stations in the Los Angeles region, from the CyberShake platform (Graves *et al.*, 2011; Jordan *et al.*, 2018). CyberShake is a high-performance computing platform developed to conduct simulation-based seismic hazard analysis. The platform simulates earthquake ruptures and resulting wave propagation through three-dimensional velocity models of the earth, including the effects of geological conditions such as deep sedimentary basins. The earthquake data are numerically simulated ground shaking intensity (measured via three second spectral acceleration, $SA(3s)$) at each station from a Magnitude 6.65 earthquake rupture on the Puente Hills fault. 32 simulations of earthquake ruptures were produced (with the epicenter and slip characteristics varied for each simulation), and a resulting ground shaking intensity is produced at each station (Maechling *et al.*, 2007). For each observed intensity at a given station i, we calculate the sample mean of (log) intensity at station i over a series of realizations with the same rupture. We then compute a within-event residual representing variation of the intensity relative to its mean. Figure 8 shows the locations of the stations and the rupture considered in this study. Figure 9 shows an example of simulated ground motion intensities in the study region.

The period of three seconds is chosen because it is sensitive to the potential effects of geologic structure, and non-stationarity of its correlations has been qualitatively observed (Chen and Baker, 2019). It is thus a useful choice to illustrate the potential of this approach to quantify non-stationarity. Because these simulations reflect the physical processes associated with earthquake rupture and wave propagation, stations with nearby locations and similar geological conditions are anticipated to experience similar shaking. We are interested in applying community detection to identify locations of communities and compare them with geographical information. Figure 10 shows depths to material with a shear-wave velocity of 1 km/s ($Z_{1.0}$) in the region, as a reference to varying geological conditions. The parameter $Z_{1.0}$ is selected because it is a popular predictor variable for basin effects on long-period SA (Chiou and Youngs, 2014; Day *et al.*, 2008).

5.2 Graph Construction

For every pair of stations separated by less than 100 km, we use Equation 1 to calculate the correlation coefficient in ground motion intensity measure within-event residual. Figure 11 shows calculated correlation coefficients for all pairs of nodes plotted versus distance. The following global exponential correlation function model is fitted to the data, as this function is commonly used in modeling spatial correlations of ground motions (Baker and Chen, 2020; Schiappapietra and Douglas, 2020).

$$\rho(d) = \exp\left(\frac{-3d}{r}\right) \tag{26}$$

Where d is the separation distance and r is the parameter to be fitted. In Figure 11, the fitted $\hat{r}$ is 79.6 km using a weighted least square approach (Baker and Chen, 2020). It is evident that the correlations generally decrease with increasing separation distance, and that the correlation of an individual station pair can differ

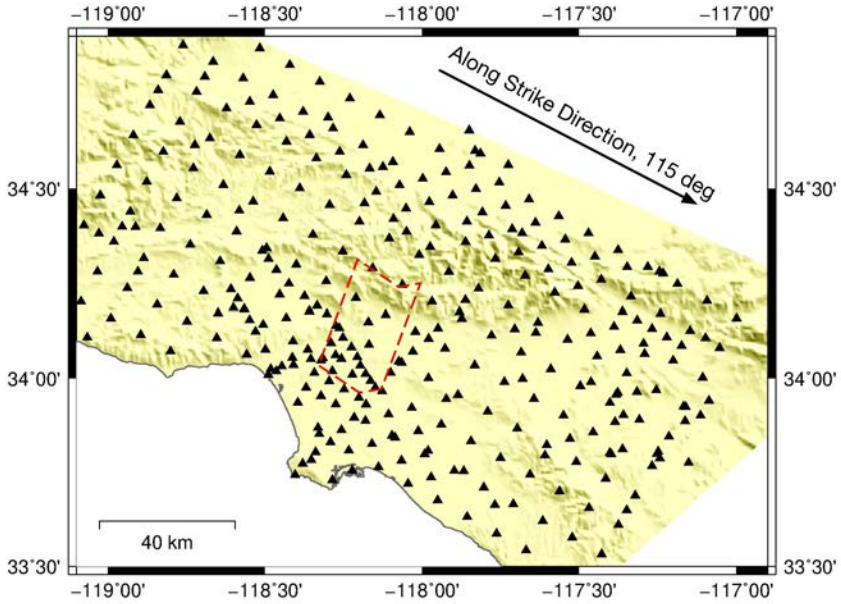


Figure 8: Locations of the stations (shown with triangles) and the surface projection of the rupture (shown with red dashed lines) in the study region. The arrow shows the azimuth for the rupture strike direction.

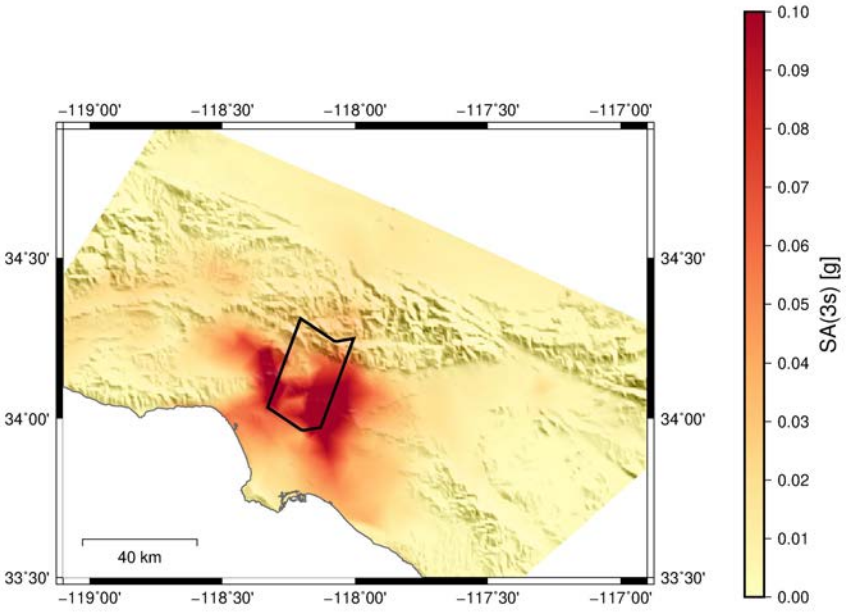


Figure 9: Example of simulated ground motion intensities in the study region from the Puente Hills rupture. Black lines show the surface projection of the rupture.

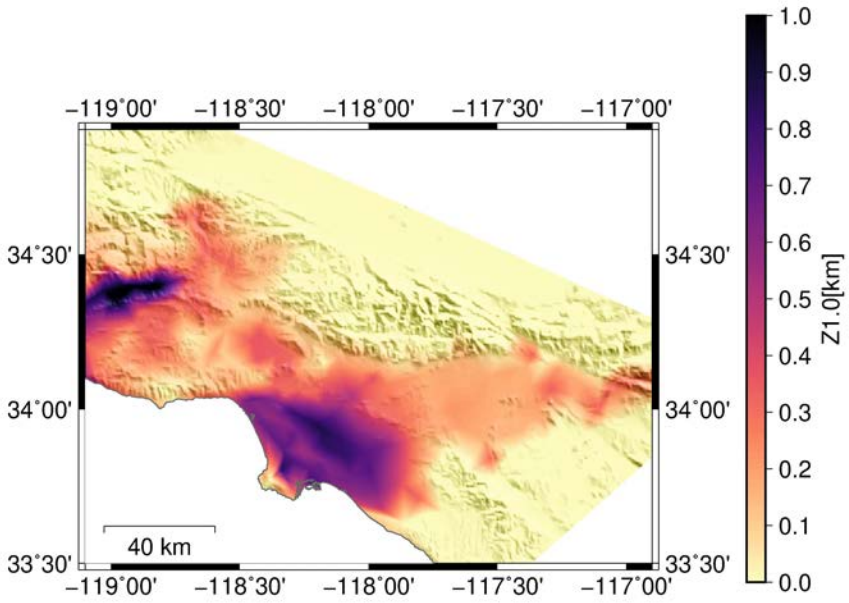


Figure 10: Map of depths to material with 1 km/s shear wave velocity ($Z_{1.0}$) in the study region. Blue regions indicate sedimentary basins.

significantly from the global correlation. We then calculate the correlation deviations using Equation 21. These deviations are the elements of the adjacency matrix of the constructed correlation deviation graph. There are 335 nodes and 42,144 edges in the graph.

Figure 12 shows the first 15 eigengaps of the constructed graph. The magnitudes of the eigengaps peak at $k = 2$ and reach an initial low point at $k = 5$, suggesting that there may be 2 to 4 main communities in the graph. The algorithm sometimes (such as in this example) clusters a single node into one community, and thus we set the number of communities equal to 5 in order to get 4 non-trivial communities. Additionally, there are somewhat larger eigengaps at $k = 9$, 12, and 15, so we also try 15 clusters as an alternative.

5.3 Community Detection Results

Figure 13 shows the community detection results for the correlation deviation graph using spectral clustering. The detected communities show consistency with their geological conditions and relative position to the rupture. In Figure 13a, when we set the number of communities to 5, the algorithm identifies 4 major communities, which correspond to the basin seen in Figure 10 (community 4) and three outside regions (communities 1, 2, and 3). Communities 1, 2, and 3 are aligned along the rupture strike direction.

Figure 13b shows results for $k = 15$. There are 10 detected communities with more than 15 stations. The communities in Figure 13a are further divided into smaller groups in this case. For example, the basin region is divided into communities 7 and 9. The boundary between communities 2 and 4 in Figure 13a is identified as community 8 in Figure 13b.

We rearrange the adjacency matrix according to the community assignment to visualize the edge weights within and across communities. In Figure 14, each pixel shows the edge weight between two nodes. We order the nodes (the rows and columns of the adjacency matrix) according to their community. For example, the upper left corner of Figure 14a shows the edge weights between the nodes in the first community in Figure

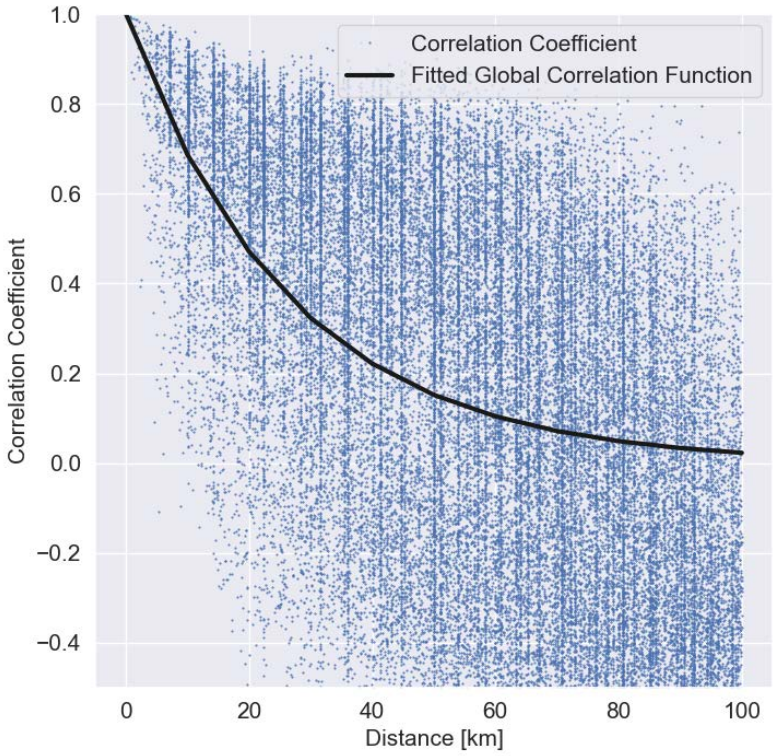


Figure 11: Correlation coefficients of all pairs of nodes as a function of separation distance.

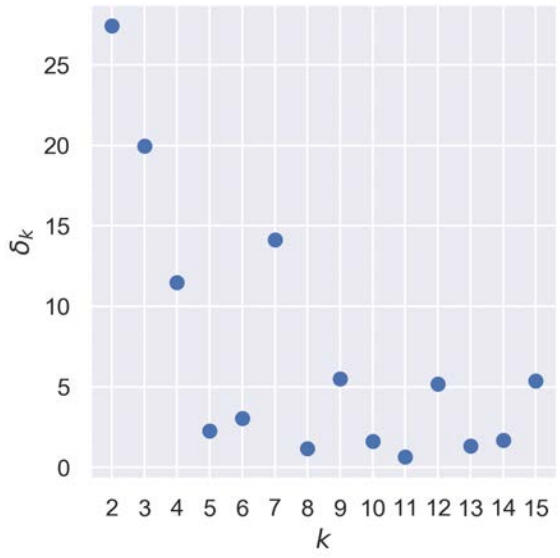
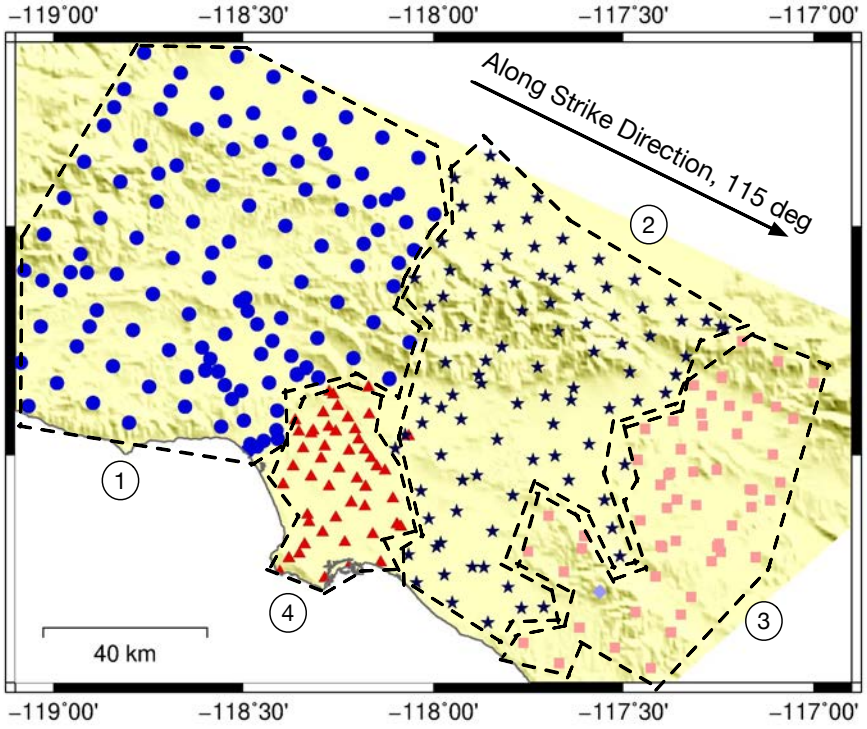
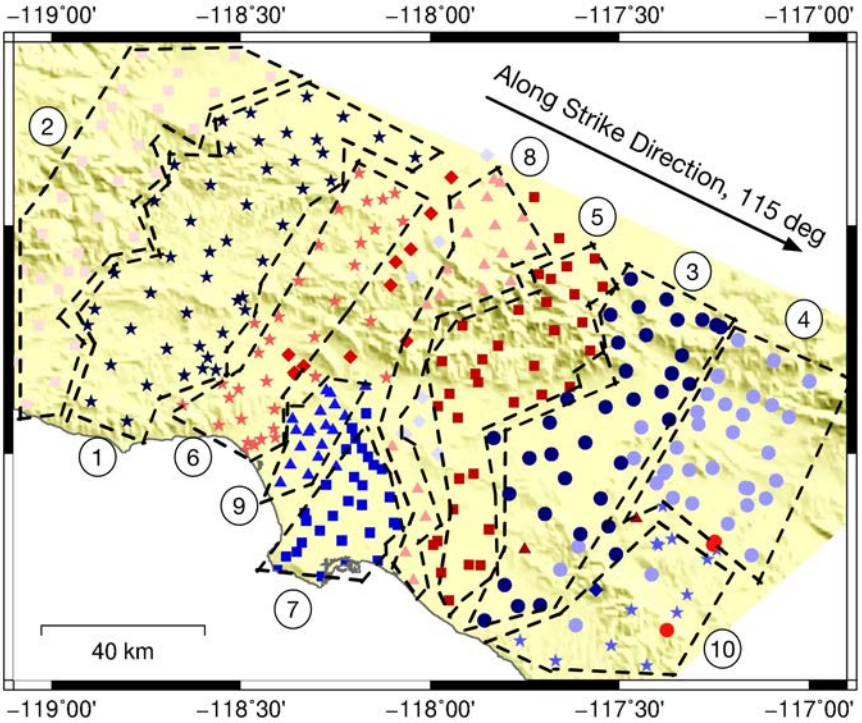


Figure 12: Eigengaps of the constructed correlation deviation graph.

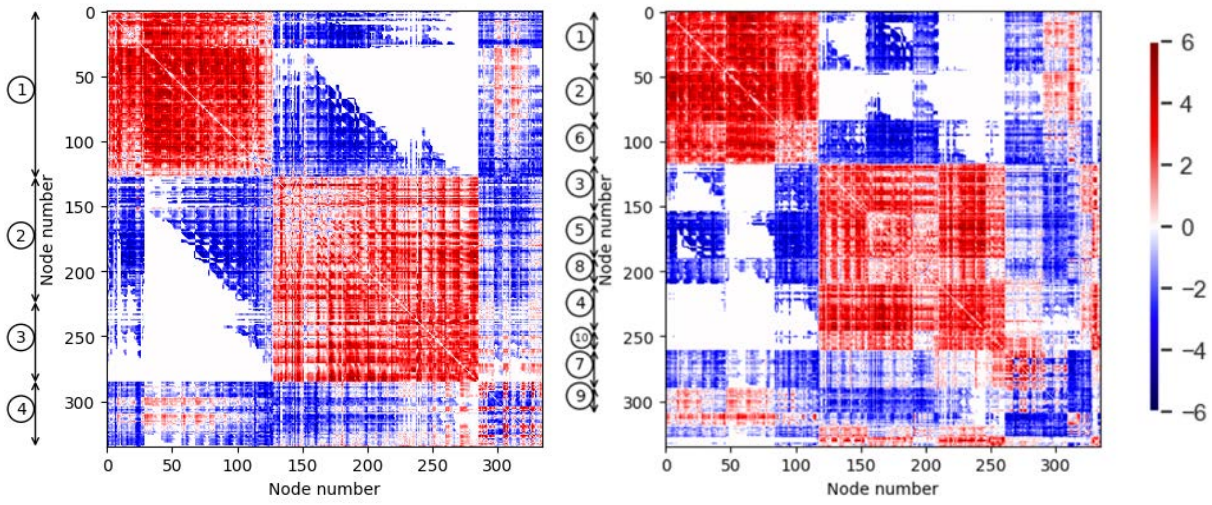


(a)



(b)

Figure 13: Detected communities using the correlation deviation graph and spectral clustering with (a) 5 communities and (b) 15 communities. Nodes of different communities are shown in different symbols. Dash lines and numbers show the communities with more than 15 stations.



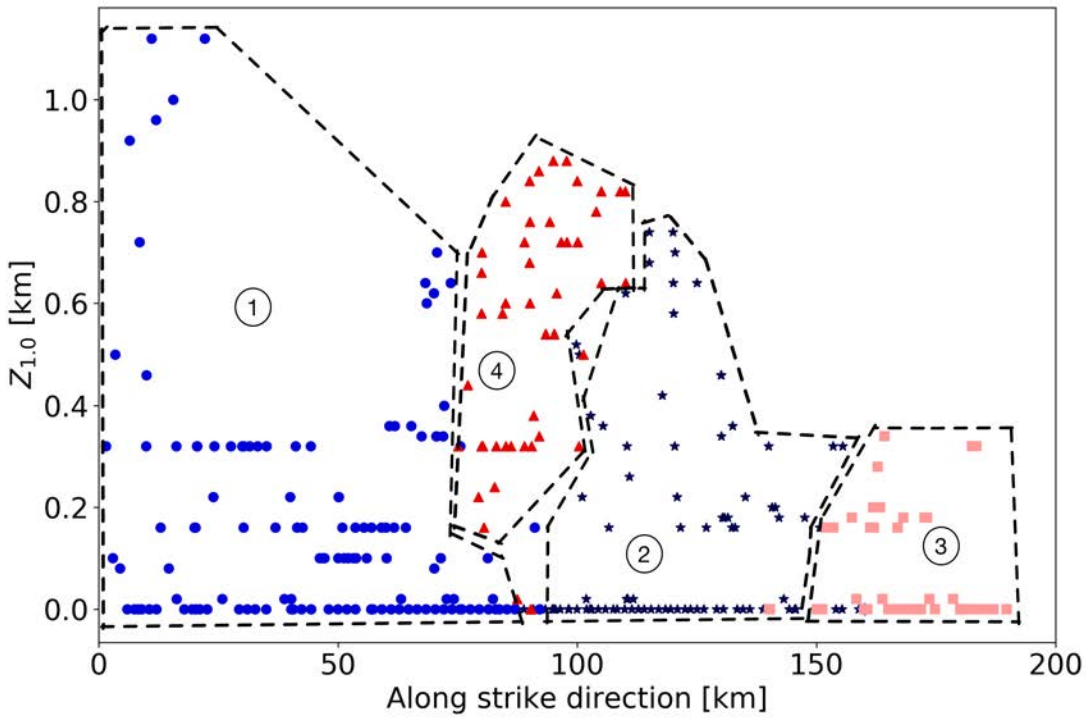
(a) **(b)**

Figure 14: (a) Graph adjacency matrix, with nodes rearranged into 5 communities. (b) Graph adjacency matrix, with nodes rearranged into 15 communities. Red indicates positive edge weight, and blue indicates negative edge weight. White pixels indicate that correlations were not computed because the nodes are separated by more than 100 km. Circled numbers at the left of each subfigure indicate the communities labelled in Figure 13.

13a. The rearranged adjacency matrix shows block patterns. It suggests that the shaking correlation is generally higher than expected if a pair of stations is in the same community (i.e., most diagonal blocks are red). At the same time, it is often lower than expected between nodes belonging to different communities (i.e., some off-diagonal blocks are blue). A notable exception is the off-diagonal block between communities 2 and 3 in Figure 14a, which is mostly red and indicates that these two communities have nodes with high correlations across the communities. This suggests that these two communities should be viewed as a single community, and so we combine them in further analysis below. Figure 14b shows the same edge weights rearranged according to the communities in Figure 13b. Compared with Figure 14a, the block pattern is similar, indicating that further subdividing the initial 4 communities did not produce more community structure. In both figures, there are three main blocks, which indicates that three is close to an optimal number of communities. This is also consistent with the eigengaps in Figure 12, as the eigengaps were largest for $2 \leq k \leq 4$.

Figure 15 shows the community assignment results plotted versus node $Z_{1.0}$ value and its relative position along the rupture strike direction. This position is defined as the position of stations projected to the rupture strike azimuth. The communities are well separated in the plot, which suggests that these metrics are two influential factors for community assignment (and thus spatial correlation). Qualitatively, we anticipate that nodes with similar geological conditions will experience greater similarity in shaking intensity due to common site effects. Further, we anticipate that nodes with similar orientation relative to the fault will experience common wave propagation effects and thus more similar shaking. This suggests that basin effects explain some non-stationarity. Additionally, much of the clustering reflects groupings away from the rupture, indicating spatial structure potentially related to wave propagation effects.

To quantitatively study the difference of spatial correlations across communities, we refit the correlation function of Equation 26 for stations within each community and across communities. Figure 16 shows the



(a)

Figure 15: Community detection results plotted versus node $Z_{1.0}$ value and its relative position along the rupture strike direction with 5 communities.

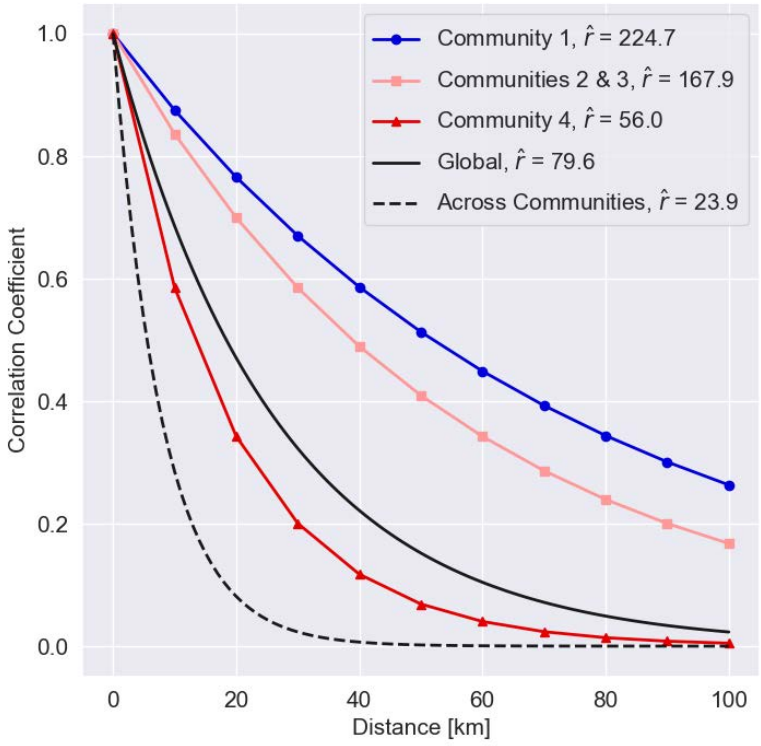


Figure 16: Refitted correlation models for different communities. Community numbers refer to the communities in Figure 13a. The Global model is from Figure 11 and the Across Communities model is fit to all pairs of stations that belong to different communities. $\hat{r}$ is the fitted parameter in Equation 26

refitted correlation models. As expected, the within community correlation is higher than across communities. However, different communities may have different correlation functions. For example, communities 1 and 2-3 are much more highly correlated, while community 4 shows similar correlation to the global correlation. This suggests that these highly correlated regions are more likely to experience spatial uniformity in ground shaking intensity, which implies higher seismic risks, as the shaking intensity in the regions is more likely to be higher at the same time. Conversely, the correlations between nodes of differing communities are notably lower than the global correlation model. These differences were previously not possible to quantitatively identify and incorporate into spatial correlation models.

6 Conclusions

We propose a community detection method for spatially correlated data. We first compute the correlation deviation for each pair of locations with respect to a global correlation model, and then construct a correlation deviation graph where the edge weights are equal to these deviations. This results in a graph with both positive and negative links, and the weight of the link is normally distributed. Signed spectral clustering is then applied to this graph to detect communities with higher correlation.

We compare the correlation deviation graph with traditional correlation graph construction approaches by testing the community detection accuracy on synthetic correlation data. For data with a global correlation in addition to community-specific correlation effect, the correlation deviation graph outperforms alternatives over a range of conditions. We also compare the performance of signed spectral clustering and the signed Louvain algorithm for community detection on these graphs, and find that the signed spectral clustering algorithm performs best. Furthermore, we show that the last significant eigengap's position can indicate the number of communities in the graph.

We then apply the proposed method to simulated earthquake ground motion data: a data set where this community detection task is of practical interest. We find that the proposed algorithm is able to detect the regions with higher correlations in ground motion intensities, and that the detected communities relate to areas with comparable geological conditions or geometry relative to the earthquake rupture. This helps analysts understand the potential underlying factors causing higher correlated shaking, and provides a quantitative pathway towards incorporating those factors in correlation models.

Furthermore, the proposed method provides a new approach for simulating ground motions. By first identifying communities, and individually fitting a correlation model for each community, the community-based correlation modeling approach has the ability to capture spatially non-stationary correlations. Compared with current global stationary models, the proposed approach has the potential to model more complex correlation structures in ground motions and thus provide refined estimates for regional seismic risks.

Acknowledgements

Thanks to Hongtao Sun and Yang Wang for helpful discussions during the early stages of this research. Thanks to Scott Callaghan for assistance accessing the CyberShake data, and Brendon Bradley and Ethan Thomson for assistance with shear wave velocity data.

Project Data

Earthquake shaking data used in this study are from the version 15.12 CyberShake runs, which can be obtained at `https://scec.usc.edu/scecpedia/Accessing_CyberShake_Database_Data` (last accessed July 2020). Figure 8, 10 and 13 were made using the Generic Mapping Tools version 5.4.4 Wessel and Smith (1998). Code used to perform the above analysis and produce the figures in this report is available at `https://github.com/yilinchen0911/correlation-deviation-graph`.

Bibliography of Publications Resulting From the Work Performed Under the Award

No publications have resulted yet from this Award.

The work described in this report is currently under review for publication as an archival journal article.

References

Alexiadis, M., P. Dokopoulos, and H. Sahsamanoglou (1999). Wind speed and power forecasting based on spatial correlation models, *IEEE Transactions on Energy Conversion* **14**(3), 836–842.

Baker, J. W., and Y. Chen (2020). Ground motion spatial correlation fitting methods and estimation uncertainty, *Earthquake Engineering & Structural Dynamics* **49**(15), 1662–1681.

Bansal, N., A. Blum, and S. Chawla (2004). Correlation clustering, *Machine learning* **56**(1-3), 89–113.

Barthélemy, M. (2011). Spatial networks, *Physics Reports* **499**(1-3), 1–101, ISSN 03701573, doi:10.1016/j.physrep.2010.11.002.

Bhowmick, S. S., and B. S. Seah (2015). Clustering and summarizing protein-protein interaction networks: A survey, *IEEE Transactions on Knowledge and Data Engineering* **28**(3), 638–658.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks, *Journal of statistical mechanics: theory and experiment* **2008**(10), P10008.

Boore, D. M., J. F. Gibbs, W. B. Joyner, J. C. Tinsley, and D. J. Ponti (2003). Estimated Ground Motion From the 1994 Northridge, California, Earthquake at the Site of the Interstate 10 and La Cienega Boulevard Bridge Collapse, West Los Angeles, California, *Bulletin of the Seismological Society of America* **93**(6), 2737–2751.

Chen, Y., and J. W. Baker (2019). Spatial correlations in cybershake physics-based ground-motion simulations, *Bulletin of the Seismological Society of America* **109**(6), 2447–2458.

Chen, Y., J. Xu, and M. Xu (2015). Finding community structure in spatially constrained complex networks, *International Journal of Geographical Information Science* **29**(6), 889–911.

Chiou, B. S.-J., and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **30**(3), 1117–1153.

Clauset, A., M. E. Newman, and C. Moore (2004). Finding community structure in very large networks, *Physical review E* **70**(6), 066111.

Day, S. M., R. Graves, J. Bielak, D. Dreger, S. Larsen, K. B. Olsen, A. Pitarka, and L. Ramirez-Guzman (2008). Model for basin effects on long-period response spectra in southern california, *Earthquake Spectra* **24**(1), 257–277.

Demaine, E. D., D. Emanuel, A. Fiat, and N. Immorlica (2006). Correlation clustering in general weighted graphs, *Theoretical Computer Science* **361**(2-3), 172–187.

Duan, L., W. N. Street, Y. Liu, and H. Lu (2014). Community detection in graphs through correlation, in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1376–1385.

Emanuel, D., and A. Fiat (2003). Correlation clustering–minimizing disagreements on arbitrary weighted graphs, in *European Symposium on Algorithms*, pp. 208–220, Springer.

Esmailian, P., and M. Jalili (2015). Community Detection in Signed Networks: The Role of Negative ties in Different Scales, *Scientific Reports* **5**, ISSN 20452322, doi:10.1038/srep14339.

Ester, M., H.-P. Kriegel, J. Sander, X. Xu, *et al.* (1996). A density-based algorithm for discovering clusters in large spatial databases with noise., in *Kdd*, volume 96, pp. 226–231.

Expert, P., T. S. Evans, V. D. Blondel, and R. Lambiotte (2011). Uncovering space-independent communities in spatial networks, *Proceedings of the National Academy of Sciences* **108**(19), 7663–7668.

Garakaninezhad, A., and M. Bastami (2017). A novel spatial correlation model based on anisotropy of earthquake ground-motion intensity, *Bulletin of the Seismological Society of America* **107**(6), 2809–2820, ISSN 19433573.

Goda, K., and G. M. Atkinson (2010). Intraevent spatial correlation of ground-motion parameters using sk-net data, *Bulletin of the Seismological Society of America* **100**(6), 3055–3067.

Goda, K., and H. Hong (2008). Estimation of seismic loss for spatially distributed buildings, *Earthquake Spectra* **24**(4), 889–910.

Gómez, S., P. Jensen, and A. Arenas (2009). Analysis of community structure in networks of correlated data, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **80**(1), ISSN 15502376, doi: 10.1103/PhysRevE.80.016114.

Goovaerts, P. (1997). *Geostatics for natural resources evaluation*, Oxford University Press.

Graves, R., T. H. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner, *et al.* (2011). Cybershake: A physics-based seismic hazard model for southern california, *Pure and Applied Geophysics* **168**(3-4), 367–381.

Heresi, P., and E. Miranda (2019). Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates, *Bulletin of Earthquake Engineering* **17**(3), 1099–1115.

Huynh, X.-H., F. Guillet, and H. Briand (2006). Evaluating interestingness measures with linear correlation graph, in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 312–321, Springer.

Jayaram, N., and J. W. Baker (2009). Correlation model for spatially distributed ground-motion intensities, *Earthquake Engineering & Structural Dynamics* **38**(15), 1687–1708.

Jayaram, N., and J. W. Baker (2010). Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment, *Earthquake Engineering & Structural Dynamics* **39**(10), 1109–1131.

Jordan, T. H., S. Callaghan, R. W. Graves, F. Wang, K. R. Milner, C. A. Goulet, P. J. Maechling, K. B. Olsen, Y. Cui, G. Juve, K. Vahi, J. Yu, E. Deelman, and D. Gill (2018). Cybershake models of seismic hazards in southern and central California, in *Proceedings of the US National Conference on Earthquake Engineering*, Earthquake Engineering Research Institute, Los Angeles, California, USA, ISBN 9781510873254.

Kanavos, A., I. Perikos, I. Hatzilygeroudis, and A. Tsakalidis (2018). Emotional community detection in social networks, *Computers & Electrical Engineering* **65**, 449–460.

Kunegis, J., S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak (2010). Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization, in *Proceedings of the 2010 SIAM International Conference on Data Mining*, pp. 559–570, doi:10.1137/1.9781611972801.49.

Lanzano, G., F. Pacor, L. Luzi, M. D'Amico, R. Puglia, and C. Felicetta (2017). Systematic source, path and site effects on ground motion variability: the case study of Northern Italy, *Bulletin of Earthquake Engineering* **15**(11), 4563–4583, ISSN 15731456.

Leskovec, J., and E. Horvitz (2014). Geospatial structure of a planetary-scale social network, *IEEE Transactions on Computational Social Systems* **1**(3), 156–163.

Loth, C., and J. W. Baker (2013). A spatial cross-correlation model of spectral accelerations at multiple periods, *Earthquake Engineering & Structural Dynamics* **42**(3), 397–417.

MacMahon, M., and D. Garlaschelli (2015). Community detection for correlation matrices, *Physical Review X* **5**(2), 021006.

Maechling, P., E. Deelman, L. Zhao, R. Graves, G. Mehta, N. Gupta, J. Mehringer, C. Kesselman, S. Callaghan, D. Okaya, *et al.* (2007). Scec cybershake workflowsâĂŤautomating probabilistic seismic hazard analysis calculations, in *Workflows for e-Science*, pp. 143–163, Springer.

Markhvida, M., L. Ceferino, and J. W. Baker (2018). Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics, *Earthquake Engineering and Structural Dynamics* **47**(5), 1107–1123, ISSN 10969845, doi:10.1002/eqe.3007.

Ng, A. Y., M. I. Jordan, and Y. Weiss (2002). On spectral clustering: Analysis and an algorithm, in *Advances in neural information processing systems*, pp. 849–856.

North, G. R., J. Wang, and M. G. Genton (2011). Correlation models for temperature fields, *Journal of Climate* **24**(22), 5850–5862.

Park, J., P. Bazzurro, and J. W. Baker (2007). Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation, *Applications of statistics and probability in civil engineering* pp. 1–8.

Pedronette, D. C. G., and R. d. S. Torres (2016). A correlation graph approach for unsupervised manifold learning in image retrieval tasks, *Neurocomputing* **208**, 66–79, ISSN 18728286, doi:10.1016/j.neucom.2016.03.081.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, *Journal of Machine Learning Technologies* **2**, 37–63.

Schiappapietra, E., and J. Douglas (2020). Modelling the spatial correlation of earthquake ground motion: Insights from the literature, data from the 2016–2017 central italy earthquake sequence and ground-motion simulations, *Earth-Science Reviews* p. 103139.

Sgobba, S., G. Lanzano, F. Pacor, R. Puglia, M. D'amico, C. Felicetta, and L. Luzi (2019). Spatial correlation model of systematic site and path effects for ground-motion fields in northern Italy, *Bulletin of the Seismological Society of America* **109**(4), 1419–1434, ISSN 19433573.

Van Gennip, Y., B. Hunter, R. Ahn, P. Elliott, K. Luh, M. Halvorson, S. Reid, M. Valasik, J. Wo, G. E. Tita, *et al.* (2013). Community detection using spectral clustering on sparse geosocial data, *SIAM Journal on Applied Mathematics* **73**(1), 67–83.

Von Luxburg, U. (2007). A tutorial on spectral clustering, *Statistics and computing* **17**(4), 395–416.

Weatherill, G., V. Silva, H. Crowley, and P. Bazzurro (2015). Exploring the impact of spatial correlations and uncertainties for portfolio analysis in probabilistic seismic loss estimation, *Bulletin of Earthquake Engineering* **13**(4), 957–981.

Wessel, P., and W. H. Smith (1998). New, improved version of Generic Mapping Tools released, *Eos, Transactions American Geophysical Union* **79**(47), 579–579.

Wu, D., J. Shi, and N. Mamoulis (2017). Density-based place clustering using geo-social network data, *IEEE Transactions on Knowledge and Data Engineering* **30**(5), 838–851.

Yao, K., D. Papadias, and S. Bakiras (2019). Density-based community detection in geo-social networks, in *Proceedings of the 16th International Symposium on Spatial and Temporal Databases*, pp. 110–119.

Yildirimoglu, M., and J. Kim (2018). Identification of communities in urban mobility networks using multi-layer graphs of network traffic, *Transportation Research Part C: Emerging Technologies* **89**, 254–267.

Yiu, M. L., and N. Mamoulis (2004). Clustering objects on a spatial network, in *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pp. 443–454.

Zelnik-Manor, L., and P. Perona (2005). Self-tuning spectral clustering, in *Advances in neural information processing systems*, pp. 1601–1608.